*Coupling vs. Conductance for the Jerrum–Sinclair Chain**

V.S. Anil Kumar, H. Ramesh
Department of Computer Science and Automation, Indian Institute of Science, Bangalore, 560012

Received 14 April 1999; revised 21 March 2000; accepted 31 May 2000

ABSTRACT: We address the following question: is the causal coupling method as strong as the conductance method in showing rapid mixing of Markov chains? A causal coupling is a coupling which uses only past and present information, but not information about the future. We answer the above question in the negative by showing that there exists a bipartite graph G such that any causal coupling argument on the Jerrum–Sinclair Markov chain for sampling almost uniformly from the set of perfect and near perfect matchings of G must necessarily take time exponential in the number of vertices in G. In contrast, the above Markov chain on G has been shown to mix in polynomial time using conductance arguments. © 2001 John Wiley & Sons, Inc. Random Struct. Alg., 18, 1–17, 2001

1. INTRODUCTION

Simulating a rapidly mixing Markov chain is a popular method to construct an almost uniform sampling scheme for an exponentially large population. A rigorous analysis of the efficiency of this method was pioneered by Broder [3] in his seminal paper on the approximation of the permanent (i.e., approximately counting the number of perfect matchings in a given bipartite graph). Since then this technique has been applied to many problems in approximate counting and sampling 14–19.

Correspondence to: H. Ramesh; e-mail: ramesh@csa.iisc.ernet.in.
*An extended abstract of this work appeared in the Proceedings of the 40th IEEE Symposium on Foundations of Computer Science, 1999.

Techniques for Showing Rapid Mixing. The main challenge in using the above technique is to bound the mixing time of the Markov chain in question. Broder used a technique called *coupling* to show that the mixing time was polynomial, provided the ratio of the number of near perfect matchings (i.e., those with all but one vertex on either side matched) to perfect matchings is polynomial. Coupling was known earlier as a method to prove that a given chain achieves the right stationary distribution [20, 24]. Aldous [1] was the first to use coupling to show rapid mixing. We will describe this technique in detail shortly. However, as pointed out by Mihail [23], there was an error in Broder's coupling argument. Jerrum and Sinclair [17] later showed that Broder's chain (actually a slight variant of it) mixes rapidly using a completely different technique, which involved showing that the underlying graph had large *conductance*, by demonstrating the existence of several *canonical paths* with low edge congestion.

The Coupling Method. The coupling argument to show that a given Markov chain mixes rapidly starting from initial distribution π' proceeds as follows.

Consider two processes $\mathscr{X}, \mathscr{Y}$ on the same state space as the above Markov chain with the following properties. $\mathscr{X}$ is a copy of the above Markov chain and starts from the stationary distribution. $\mathscr{Y}$ starts from the distribution π' and makes its moves dependent on the moves of $\mathscr{X}$. However, $\mathscr{Y}$ by itself is also a faithful copy of the original chain, i.e., for all pairs of states a, b, the transition probability of $\mathscr{Y}$ from a to b, given it is in state a, is identical to the corresponding probability for the original Markov chain. Thus, if $\mathscr{X}_t, \mathscr{Y}_t$ denote the respective states at time instant t, then $\mathscr{Y}_t$ could be dependent on $\mathscr{X}_0 \cdots \mathscr{X}_t$ and on $\mathscr{Y}_0 \cdots \mathscr{Y}_{t-1}$, which could make the joint process $(\mathscr{X}, \mathscr{Y})$ non-Markovian, as long as $\mathscr{Y}$ remains faithful to the original chain. This is formally described in Section 2A.

The random variable of interest is the *coupling time*, which is the least time instant t at which $\mathscr{X}_t = \mathscr{Y}_t$. Clearly, once $\mathscr{X}_t = \mathscr{Y}_t$, the two processes can make identical moves subsequently and always be in identical states. The coupling lemma [1, 15] states that the probability that the coupling time exceeds some value t is an upper bound on the variation distance between the stationary distribution π of the Markov chain under consideration and the distribution of this chain at time t starting from distribution π'. Therefore, to show rapid mixing, it suffices to exhibit a process $\mathscr{Y}$ with the properties of the above paragraph and then show that the coupling time is small with high probability.

The above definition of coupling can in fact be generalized to allow for $\mathscr{Y}$ to make its moves dependent on future moves of $\mathscr{X}$, in addition to the past moves of $\mathscr{X}, \mathscr{Y}$, i.e., $\mathscr{Y}_t$ can now depend upon $\mathscr{X}_{t+1}, \mathscr{X}_{t+2}$, etc. As long as the distribution of $\mathscr{Y}$ after t' steps is identical to the distribution of the original Markov chain after t' steps (starting from π'), the coupling lemma continues to hold for the value t'; i.e., the probability that the coupling time exceeds t' is an upper bound on the variation distance between the stationary distribution π of the Markov chain under consideration and the distribution of this chain at time t' starting from distribution π'.

We call couplings which looks into the future *noncausal*, as opposed to *causal* couplings which only use present and past information; i.e., $\mathscr{Y}_t$ depends only on $\mathscr{X}_0 \cdots \mathscr{X}_t, \mathscr{Y}_0 \cdots \mathscr{Y}_{t-1}$.

The coupling technique is described in detail in [15, 20]; the former deals only with causal couplings. Griffeath [12] showed that there always exists a maximal coupling, i.e., a noncausal coupling whose coupling time equals the mixing time of the Markov chain. Thus, in principle, it is possible to show rapid mixing (if the chain in question indeed mixes rapidly) by setting up an appropriate noncausal coupling and bounding its coupling time. However, the latter task is typically hard. In this paper, we study only causal couplings. We would like to mention that though Markovian couplings seem less general than causal couplings, in fact, there is always a Markovian coupling that has almost the same expected coupling time as any causal coupling. We do not know whether this is true of the coupling time as well.

Coupling vs. Conductance. Jerrum and Sinclair [17] showed that a Markov chain mixes rapidly if and only if it has large conductance. Our original aim was to study the following question: is the coupling method as powerful as the conductance method for showing rapid mixing? In other words, if the Markov chain in question has large conductance, then does it also admit a causal coupling argument, up to polynomial factors? Or can the admissibility of causal coupling arguments be characterized in terms of some properties of the underlying graph?

There are many situations where coupling has been used to show rapid mixing. Some examples are Markov chains related to the following tasks: estimating the volume of convex bodies [6], counting the number of linear extensions of a partial order [5], counting k-colourings [4, 14], and counting independent sets [22]. There are others in which conductance has been used, e.g., in Markov chains related to the following tasks: estimating the volume of convex bodies [10, 21], counting matchings in dense graphs [17], and estimating the partition function in the Ising model [18].

To the best of our knowledge, most instances of coupling in the theoretical computer science literature seem to actually be causal couplings.[1] The one exception we know of is the paper by Czumaj et al. [8], which uses a noncausal coupling to show rapid mixing for a chain on permutations. One reason for the popularity of causal couplings is the relative ease of setting up and analyzing causal couplings over noncausal ones.

Mihail [23] and Sinclair [25] point out that coupling seems ill-suited to prove rapid mixing for chains which lack symmetry. Recent work by Burdzy and Kendall [7] gave the first formal result on the weakness of causal couplings. They showed the existence of chains for which no *efficient* causal coupling exists; i.e., there is no causal coupling which couples at the maximum possible exponential rate, given by the spectral gap. However, this does not seem to preclude the existence of causal couplings which couple at "approximately" this maximum rate. On the other hand, there have been recent results [22] where coupling has been used in situations in which the underlying Markov chain lacks symmetry.

Our Result. We show that there exists a bipartite graph G with the following property: for any causal coupling argument on the Jerrum–Sinclair chain [17] for sampling perfect and near perfect matchings in G, the probability that the coupling time is $O(c^n(1-\epsilon))$ is more than ϵ, for some constant $c > 1$. The ratio of number of near perfect matchings to the number of perfect matchings in G will be polynomial,

[1] The coupling used by Broder [3] was non-Markovian in that it made moves which are functions of past history, but nevertheless causal.

so that the above chain can be shown to mix in polynomial time using conductance arguments.

The bipartite graph G above will have n vertices on either side along with some additional properties, which will help prove lower bounds on the coupling time. Consider any causal coupling process [i.e., the process $(\mathcal{X}_t, \mathcal{Y}_t)$] for the Jerrum–Sinclair chain on graph G. The *distance* $d(\mathcal{X}_t, \mathcal{Y}_t)$ is defined to be the size of the symmetric difference between these two perfect or near perfect matchings. We show the following.

(1) If $\mathcal{X}$, $\mathcal{Y}$ start from some two fixed states $\mathcal{X}_0, \mathcal{Y}_0$, respectively, then the probability that the coupling time exceeds t is at least ϵ, for all t smaller than $1 - \epsilon$ times some exponential in $d(\mathcal{X}_0, \mathcal{Y}_0)$. In addition, if $\mathcal{X}$ starts in π and $\mathcal{Y}$ starts from any arbitrary distribution π', then most of the probability mass will indeed be on pairs of starting states having distance $\Theta(n)$.

(2) The expected coupling time is exponential in n, *irrespective* of the distance between the starting states (provided, of course, that the distance is nonzero).

Our proof has two main aspects.

Bounding Transition Probabilities. The key observation we make is that the coupling process $(\mathcal{X}_t, \mathcal{Y}_t)$ behaves essentially like a linear chain; i.e., the probability that $d(\mathcal{X}_t, \mathcal{Y}_t) < d(\mathcal{X}_{t-1}, \mathcal{Y}_{t-1})$ for fixed $\mathcal{X}_t, \mathcal{Y}_t$ is only a constant fraction of the probability that $d(\mathcal{X}_t, \mathcal{Y}_t) > d(\mathcal{X}_{t-1}, \mathcal{Y}_{t-1})$. Showing this involves bounding the probabilities of *distance increasing* and *distance decreasing* transitions for any coupling strategy. We show lower bounds on the probabilities of distance increasing transitions and upper bounds on the probabilities of distance decreasing transitions from every pair of states $\mathcal{X}_t, \mathcal{Y}_t$ for which $d(\mathcal{X}_t, \mathcal{Y}_t)$ is not too large. Certain properties of graph G will be used critically in this process. An important point to be noted is that these bounds need to hold for every coupling strategy; therefore we are constrained to using only the fact that $\mathcal{X}$ and $\mathcal{Y}$ must, individually, be faithful copies of the Jerrum–Sinclair chain, and not the exact nature of the coupling strategy, in deriving them.

Expectation and Tail Bounds for the Coupling Time. To bound the coupling time, we use a submartingale inequality derived along the lines of the inequalities in [13]. This inequality essentially states that if the expected change in distance at each step is a positive constant (i.e., the distance increases) and the maximum absolute value change in distance at each step is bounded by another constant, then (1) and (2) above hold. An example of such an analysis appears in a paper by Sasaki and Hajek [26]. Their process is layered, with transitions only between adjacent layers. Our process is more general with transitions going across a constant number of layers; in addition, the expected change in distance is a positive constant only if the current distance is not too large. The submartingale inequality we derive continues to hold for these situations.

Road-Map. Section 2 gives some preliminary definitions, a formal definition of the causal coupling method, and a description of the Jerrum–Sinclair chain. Section 3 shows the existence of a graph G with certain properties for which we claim that any coupling strategy on Jerrum and Sinclair's chain will require time exponential in n. In Section 4, we will set up the coupling process and bound the probabilities of distance modifying transitions in this process. Section 5 will describe a submartingale inequality which will be useful in proving bounds on the coupling time of the above process. Finally, Section 6 will describe the bounding of coupling time.

2. PRELIMINARIES

Let G denote any bipartite graph with $2n$ vertices and let V_1, V_2 denote the left and the right vertex sets, respectively, each having n vertices. Let m denote the number of edges in G.

For a process $\mathcal{X}$, let $p_{\mathcal{X}}(a, b)$ denote the probability of the transition from state a to state b, given that the current state is a.

A. The Causal Coupling Method

All further references to coupling will actually be to causal coupling. Suppose we wish to show that $\mathcal{X}$ converges to its stationary distribution π starting at distribution π'.

The Coupling Process $\mathcal{M}$. We take another copy $\mathcal{Y}$ of $\mathcal{X}$, and consider a process $\mathcal{M}$ whose state space is the Cartesian product of the state spaces of $\mathcal{X}$ and $\mathcal{Y}$. This process is defined as follows. The initial distribution for the states of $\mathcal{M}$ is just $\pi' \times \pi$. Moves in $\mathcal{M}$ from one joint state to another are made according to probabilities satisfying the following conditions.

Transitions in $\mathcal{M}$. For states of the form (a, a) in $\mathcal{M}$, $p_{\mathcal{M}}((a, a), (b, b))$ equals $p_{\mathcal{X}}(a, b)$. The transition probabilities from any other state $v = (x, y)$, $x \neq y$ satisfy the conditions mentioned below. It is important to note that $p_{\mathcal{M}}(v, w)$ could be a function of the history; i.e., the transition probabilities could vary with time. We do not show the dependence on time for notational convenience, but it should be treated as implicit.

1. For each $x' \in \mathcal{X}$ and for each time instant t, $\sum_{w \in T(x')} p_{\mathcal{M}}(v, w) = p_{\mathcal{X}}(x, x')$, where $T(x')$ is the set of states in $\mathcal{M}$ whose first entry is x'.
2. For each $y' \in \mathcal{Y}$ and for each time instant t, $\sum_{w \in T(y')} p_{\mathcal{M}}(v, w) = p_{\mathcal{Y}}(y, y')$, where $T(y')$ is the set of states in $\mathcal{M}$ whose second entry is y'.

Coupling Strategies. A coupling strategy entails specification of the probabilities $p_{\mathcal{M}}(*, *)$ satisfying the above conditions.

The Coupling Lemma. The *coupling time* is defined as the earliest instant at which the two components of the state of $\mathcal{M}$ become identical. Define t_ϵ to be the smallest integer such that the probability that the coupling time exceeds t_ϵ is below ϵ. The *coupling lemma* [9, 1, 15] states that t_ϵ is an upper bound on the time required for the chain $\mathscr{X}$ to become ϵ-close to π (which means the variation distance between the distribution of the chain and π becomes less than ϵ) starting from the initial distribution π'. We will lower bound t_ϵ by $\Omega(c^n(1-\epsilon))$, for some constant $c > 1$.

B. The Jerrum–Sinclair Chain

The states in this chain are the perfect and near perfect matchings in a bipartite graph $G = (V_1, V_2, E)$. At each state M, this chain does nothing with probability 1/2 and chooses an edge e in E uniformly at random with probability 1/2. The following action is then performed in the latter case.

1. If $e \in M$ and M is a perfect matching, e is deleted.
2. Suppose M is a near perfect matching. Let $e = (u, v)$. There are two cases.

    2.1. If u, v are both unmatched in M, then e is added.
    2.2. If exactly one of u, v is unmatched, then e is added and e' is deleted, where e' is the edge in M incident on whichever of u, v is matched; we call this a *swap* move.

3. If none of the above conditions holds, then no transition is made.

3. EXISTENCE OF GRAPH G

We will require the graph G to have the following properties. Below, an $n \times n$ bipartite graph denotes a bipartite graph with n vertices on each side.

Lemma 3.1. *There exists an $n \times n$ bipartite graph G satisfying the following properties.*

1. *The ratio of the number of near perfect matchings to perfect matchings is polynomial in n and there are $\Omega(\frac{n!}{c^n})$ perfect matchings, for some constant $c > 1$.*
2. *Each vertex has degree at least αn, for some constant $\alpha < 1/2$.*
3. *For every pair of vertices, the intersection of their neighborhoods has size at most $\alpha n/2$.*

Proof. Consider a random graph with edges chosen with probability p to be fixed shortly. Jerrum (Theorem 3.14 in [25]) showed that the probability that the ratio of the number of near perfect to perfect matchings is polynomial in n is at least is $1 - O(\frac{1}{n})$. The probability that a given vertex has fewer than $(1-\epsilon)np$ edges is at most $e^{-\epsilon^2 np/2}$. The probability that a given pair of vertices has a common neighborhood of more than $(1+\delta)np^2$, is at most $e^{-\delta^2 np^2/2}$. Both the above bounds follow from a simple application of the Chernoff bound. The total probability of one of these conditions getting violated is thus at most $O(\frac{1}{n}) + ne^{-\epsilon^2 np/2} + n^2 e^{-\delta^2 np^2/2}$; this can be made $O(\frac{1}{n})$, if we choose p, ϵ, δ to be constants so that $(1-\epsilon)p = \alpha$ and $(1+\delta)p^2 < \alpha/2$. It remains to bound the probability that the number of perfect matchings is small.

Jerrum [16] showed that the variance in the number of perfect matchings over all $n \times n$ bipartite graphs with m edges is $O(\frac{n^3}{m^2})$ times the square of the expected number of such matchings. Jerrum also showed that this expectation is $\Omega(\frac{n!}{c^n})$, if $m \geq \frac{n^2}{c}$. It follows from Chebyshev's inequality that a $1 - O(\frac{1}{n})$ fraction of all $n \times n$ bipartite graphs with m edges have $\Omega(\frac{n!}{c^n})$ perfect matchings, if $m \geq \frac{n^2}{c}$. The random graph generated in the previous paragraph has at least $\frac{n^2}{c}$ edges with probability at least $1 - e^{-\theta(n^2)}$, for some appropriate constant $c > 1$. It is now easy to see that with probability $1 - O(\frac{1}{n})$, this random graph has $\Omega(\frac{n!}{c^n})$ perfect matchings, along with the other conditions in the previous paragraph. ■

4. THE COUPLING PROCESS $\mathcal{M}$ FOR GRAPH G

The Coupling Process. Consider a coupling process $\mathcal{M}$ on the set of tuples (x, y) of perfect and near perfect matchings of G with a transition from (x, y) to (x', y') if $x \to x'$ and $y \to y'$ are transitions in the Jerrum–Sinclair chain. The probabilities for the transitions are determined by the specific coupling strategy used.

Partitioning the States. The states of the above coupling chain are partitioned into layers $L(i), i = 0 \cdots 2n$, where $L(i)$ contains all (x, y) such that $|x \oplus y| = i$ ($\oplus$ denotes the symmetric difference). Each set $L(i)$ is further partitioned into two sets $Top(i)$ and $Bot(i)$, where $Bot(i) = \{(x, y) | \exists$ vertex $v \in G$ which is unmatched in exactly one of $x, y\}$ and $Top(i) = \{(x, y)|$ either both x and y are perfect matchings or both are near perfect matchings with the same unmatched vertices$\}$.

A move in $\mathcal{M}$ from $L(i)$ to $L(j)$, is called *leftwards* or *distance reducing* if $j < i$, and *rightwards* or *distance increasing* if $j > i$. We assume that all transitions from vertices in $L(0)$ are to other vertices in $L(0)$.

Initial Distribution. Note that for any perfect or near perfect matching M, the number of perfect and near perfect matchings with symmetric difference at most $n/4$ (with respect to M) is bounded by $n\binom{n}{n/4}(\frac{n}{4})!$, which is a negligible fraction of $\frac{n!}{c^n}$, the number of perfect matchings in G (see Lemma 3.1). Thus almost all ($1 - c'^n$, for some constant $c' < 1$) of the probability mass of the initial distribution $\pi' \times \pi$ will lie on vertices in $L(i)$, $i \geq n/4$. Here π' is the initial distribution for the Jerrum–Sinclair chain on graph G and π is its stationary distribution. For simplicity, we assume that the chain $\mathcal{M}$ begins at some state in $L(i)$, $i \geq n/4$.

A. Bounds of Transition Probabilities in $\mathcal{M}$

We bound the transition probabilities from each state. Ideally, we would like each state to have small leftward probability and large rightward probability. The leftward probability from every state is indeed small. However, note that small leftward probability does not automatically imply large rightward probability as there could be transitions within the same layer. Therefore, it is not sufficient to just show that the leftward probability is small. As we shall show in Lemma 4.2 and Lemma 4.3, vertices in $Bot()$ have small leftward probability and large rightward probability for any coupling strategy, as desired. Unfortunately, for vertices in $Top()$, there are coupling strategies which ensure low or zero rightward probability. However, as we show in Lemma 4.4 and Lemma 4.5, the transition probabilities from vertices in $Bot()$ to vertices in $Top()$ are small in any coupling strategy; further, from vertices in $Top()$, there are transitions only to vertices in $Bot()$. This effectively ensures that the lack of large rightward probabilities at vertices in $Top()$ does not translate to large leftward probabilities on the whole.

There is one additional caveat. Lemmas 4.2 and 4.3 show that leftward probabilities of vertices in $Bot(i)$ are small and rightward probabilities are large, only as long as the symmetric difference i is some small enough constant fraction of n. This needs to be kept in mind for the subsequent analysis.

Lemma 4.1. *No transition in $\mathcal{M}$ can change the distance by more than* 4.

Proof. Each transition involves at most 4 edges. ∎

Lemma 4.2. *For any coupling strategy, the sum of the transition probabilities from $(x, y) \in Bot(i)$ to vertices in $L(j)$, $j < i$, is at most $\frac{2i+1}{m}$.*

Proof. Without loss of generality, assume x is a near perfect matching. A move by x is distance reducing only if one of the following holds.

1. Suppose the move by x is a swap move. Then the edge swapped out of x must not be in $x \cap y$. Indeed if the edge which is swapped out of x is in $x \cap y$ then the move by x will increase the distance by 2 and then the move by y can at most decrease the distance back by 2. The number of choices of the edge swapped in so that the edge swapped out is not in $x \cap y$ is at most $2|x - y| \leq 2i$ ($x - y$ is the set-theoretic difference; also recall that G is a bipartite graph). Therefore, the probability of such an edge being chosen is at most $\frac{2i}{2m}$ in any coupling strategy.
2. Suppose the move by x takes it to a perfect matching. In this case, the distance may or may not decrease depending upon what move y makes. But the total probability of x moving to a perfect matching is at most $\frac{1}{2m}$ in any coupling strategy.

Thus, the probability that x moves and the distance reduces is at most $\frac{2i+1}{2m}$. Next, if x does not move then there are two cases.

First, suppose y is a near perfect matching. Then the probability that y moves and the distance decreases is at most $\frac{2i+1}{2m}$ (the proof is similar to the one above for x). Second, if y is a perfect matching, then the probability that y moves and the distance decreases is at most $\frac{|y-x|}{2m} \leq \frac{i}{2m}$ (y must choose an edge not in $x \cap y$).

It follows that the total probability of distance reduction is at most $\frac{2(2i+1)}{2m} = \frac{(2i+1)}{m}$. ∎

Lemma 4.3. *For any coupling strategy, the sum of the transition probabilities from $(x, y) \in Bot(i)$ to tuples in $Bot(i+1) \cup Bot(i+2) \cup Bot(i+3) \cup Bot(i+4)$ is at least $\frac{\alpha n/2-i-2}{2m}$.*

Proof. Since $(x, y) \in Bot(i)$, one of the following conditions holds:

1. x is a near perfect matching and y is a perfect matching.
2. x is a perfect matching and y is a near perfect matching.
3. x and y are near perfect matchings with at most one common unmatched vertex.

We consider the first and the third cases in turn. The second case is symmetric to the first.

x is near perfect, y is perfect. Let $a \in V_1$ and $b \in V_2$ be the unmatched vertices in x. We consider just one situation in which the distance increases and the resulting tuple (x', y') is also in $Bot(j)$, $j > i$. This situation is when x picks edge $e = (a, u)$, where $(u, u') \in x \cap y$ and moves to $x' = x + e - (u, u')$. The distance between x and y increases by 2 in the process; i.e., $|x' \oplus y| = |x \oplus y| + 2$. y can move to y' either by doing nothing or by deleting some edge, which still results in a net increase of distance by at least 1. Further, u' and b are the unmatched vertices in x' and since $(u', b) \notin y$, at least one of them is matched in y', implying that $(x', y') \in Bot(j)$, $j > i$.

The number of edges $e = (a, u)$, where $(u, u') \in x \cap y$ is at least $\alpha n - |x - y|$. Therefore, probability of choosing such edges is at least $\frac{\alpha n - |x-y|}{2m} \geq \frac{\alpha n - |x \oplus y|}{2m} \geq \frac{\alpha n - i}{2m}$, for any coupling strategy.

x, y are both near perfect. Suppose x has vertices $a \in V_1$ and $b \in V_2$ unmatched and y has vertices $c \in V_1$ and $d \in V_2$ unmatched. Without loss of generality, assume that b and d are distinct while c could be the same as a.

We will look at just one class of moves for x, which will increase distance by 2. These moves will occur with probability at least $\frac{\alpha n/2 - |x \oplus y|}{2m}$. Then, we will show that at most two choices of moves for y can reduce the distance if x makes a move in the above class; further, these choices will reduce the distance by 2, making the net change in distance 0, and each will occur with probability at most $\frac{1}{2m}$. In addition, we will show that if x makes a move in the above class and y does not make any of the above two possible moves then the resulting tuple (x', y') will be in $Bot(j)$, $j > i$. Thus the net probability of a move to $Bot(j)$, $j > i$, is at least $\frac{\alpha n/2 - |x \oplus y|}{2m} - \frac{2}{2m} = \frac{\alpha n/2 - i - 2}{2m}$, for any coupling strategy. Note that in calculating the probability that x makes a move in the above class and y does not make any of the above two possible moves, we have not assumed anything about the probability distribution of the joint move.

The above class of moves for x involves choosing an edge $e = (b, u)$, where u is not adjacent to d and (u, u') is a matching edge in both x and y, for some $u' \in V_2$. By Lemma 3.1, there are at least $\alpha n/2$ vertices in V_1 which are adjacent to b but not to d. Thus, there are at least $\alpha n/2 - |x - y| \geq \alpha n/2 - |x \oplus y|$ edges which are in $x \cap y$ and whose endpoint in V_1 is adjacent to b but not to d. Therefore, the above choice of moves occurs with probability at least $\frac{\alpha n/2 - |x \oplus y|}{2m}$ and results in an increase in distance by 2. The only moves for y which could decrease the distance back by 2 are when it chooses the unique edge $(c, c') \in x$, if any, or the unique edge $(d, d') \in x$. If neither of these happens, then y either does nothing or swaps in an edge which is not in x; the distance cannot decrease back by 2 in either case. The probability of y choosing (c, c') or (d, d') is $\frac{2}{2m}$, for any coupling strategy. Finally,

$u' \in V_2$ is unmatched in x' and must be matched in y', as $(u, u') \in y$ and (d, u) is not an edge in G; therefore (x', y') is in $Bot(j)$, $j > i$. ■

Lemma 4.4. *For any coupling strategy, the sum of the transition probabilities from $(x, y) \in Bot(i)$ to tuples in $Top(i), Top(i+1)$ is at most $\frac{i+3}{2m}$.*

Proof. We enumerate all possibilities of moving to $Top(j)$, $j = i, i+1$, from (x, y) and verify the claim. Without loss of generality, assume x is a near perfect matching. There are two cases. The state resulting from this move is either a pair of perfect matchings or a pair of near perfect matchings with identical unmatched vertices. The former possibility happens with probability at most $\frac{1}{2m}$ for any coupling strategy, because exactly one edge can make x a perfect matching. Now we consider the latter possibility. There are two cases.

First, consider the case when x is a near perfect matching with vertices a and b unmatched and y is a perfect matching. When x moves to another near perfect matching x', one of a, b remains unmatched. Then y must delete either the matching edge incident on a or that incident on b to achieve the same pair of unmatched vertices as in x'; this happens with probability at most $\frac{2}{2m}$, for any coupling strategy.

Second, consider the case when x is a near perfect matching with vertices a and b unmatched and y is a near perfect matching with vertices c and d unmatched, $b \neq d$. We show that if $a \neq c$ then there are exactly two swap moves for x which may lead to the unmatched vertices in x', y' being identical. The probability of these moves is at most $\frac{2}{2m}$. We also show that if $a = c$, then there are at most $|x \oplus y|$ moves for x which may lead to the unmatched vertices in x', y' being identical and (x', y') being in $Top(i) \cup Top(i+1)$ [actually, (x', y') will always be in $Top(i)$ for these moves]. The probability of these moves is at most $\frac{|x \oplus y|}{2m} = \frac{i}{2m}$.

Suppose $a \neq c$. Then x must choose either the edge (a, c'), if any, such that $(c, c') \in x$, or the edge (d', b), if any, such that $(d', d) \in x$. The first leads to b, c being unmatched and the second to a, d being unmatched. These are the only pairs that can be unmatched in x' if the unmatched vertices in x', y' are identical.

Next, suppose $a = c$. Since x' and y' have to have the same unmatched vertices, $a$must either be matched in both of x', y' or be unmatched in both of them. In fact, a must be unmatched in x' and y'; otherwise b would be unmatched in x' and d in y' and since $b \neq d$, (x', y') will not be in $Top(j)$ for any j. Let g denote the other common unmatched vertex in x', y'. Let (g', g) be the unique edge in x incident on g. We claim that (g', g) must be in $x - y$, for if $(g', g) \in x \cap y$ then x must choose (g', b) and y must choose (g', d) so that the unmatched vertices become identical, but this causes the distance to increase by 2, and (x', y') would end up in $Top(i+2)$ instead of $Top(i) \cup Top(i+1)$. Thus, there are only $|x - y| \leq |x \oplus y|$ moves for x. ■

Lemma 4.5. *For any coupling strategy, the sum of the transition probabilities from $(x, y) \in Top(i)$ to tuples in $Top(j)$ is 0 for all $j \neq i$. Further, the sum of the transition probabilities to tuples in $Bot(j)$ is 0 for all $j < i - 2$.*

Proof. First, suppose that x and y are both perfect matchings. There are two possibilities: either only one of x, y deletes an edge or both x and y delete an edge each. If only one of x, y deletes an edge, the resulting tuple is in $Bot(i-1)$ or $Bot(i+1)$.

If both x and y delete the same edge, the distance remains unchanged and (x', y') remains in $Top(i)$. And if x and y delete different edges, the distance changes by at most two and the resulting unmatched vertices in x', y' are not identical. Therefore, (x', y') is in one of $Bot(i-2), \ldots, Bot(i+2)$ in this case.

Second, suppose both x and y are near perfect matchings with unmatched vertices $a \in V_1$ and $b \in V_2$. The possible moves are:

1. Both x and y add (a, b): the distance remains unchanged and $(x', y') \in Top(i)$.
2. x adds edge (a, b) and y does not add (a, b): x becomes a perfect matching while y remains a near perfect matching. y could, of course, make a swap move. But no matter what swap move it makes, the distance must always increase either by 1 or by 3. Therefore $(x', y') \in Bot(i+1) \cup Bot(i+3)$. The case when y adds edge (a, b) and x does not is symmetric.
3. x swaps in an edge $(a, c), c \neq b$, and y does not move: the distance changes by at most 2 and the unmatched vertices in x', y' are not identical. The case when y swaps in such an edge and x does not move is symmetric.
4. Both x and y swap in the same edge $(a, c), c \neq b$: let (c_1, c) be the matching edge in x and (c_2, c) be the matching edge in y. If $c_1 = c_2$, the distance is unchanged and (x', y') remains in $Top(i)$. And if $c_1 \neq c_2$, the distance reduces by two but (x', y') is in $Bot(i-2)$ because x' has c_1, b unmatched while y' has c_2, b unmatched. The case when both x and y swap in the same edge $(c, b), c \neq a$, is symmetric.
5. x and y pick distinct edges $(a, c), (a, d), c \neq d$, respectively: let (c', c) be the matching edge in x and (d', d) be the matching edge in y. If $c' = d'$, the distance does not change and (x', y') is in $Top(i)$. And if $c' \neq d'$, then the unmatched vertices in x', y' are not identical and the distance cannot decrease; i.e., $(x', y') \in Bot(i) \cup \cdots \cup Bot(i+4)$. The case when x and y pick distinct edges $(c, b), (d, b), c \neq d$, respectively, is symmetric.
6. x picks $(a, c), c \neq b$ and y picks $(d, b), d \neq a$: $(x', y') \in Bot(j)$ for some j because a is unmatched in y' but not in x'. The distance cannot reduce in the process since $(a, c), (d, b)$ are not present in $x \cup y$. The case when y picks $(a, c), c \neq b$ and x picks $(d, b), d \neq a$, is symmetric. ■

The above lemmas show that $\mathcal{M}$ has an almost linear structure with a sink and a drift away from the sink. It is possible to do a comparison argument by reducing $\mathcal{M}$ to a random walk on a line (with a sink on one end), without increasing the coupling time, and then showing that in the presence of a drift away from the sink, a random walk on the line takes a long time to reach the sink. However, we shall use martingales to obtain a more compact proof of the lower bound on the coupling time.

5. A SUBMARTINGALE INEQUALITY

We first derive a useful theorem to bound the coupling time, along the lines of the inequalities in [13]. We use the notation $E_X()$ to denote expectation with respect to a random variable X.

Theorem 5.1. *Let $X_0, X_1, X_2, \ldots$ be a sequence of random variables with the following properties (for some $R, \Delta, M > 0$):*

1. $X_i \geq 0, \forall i, i \geq 0$. *Further* $X_i = 0 \Rightarrow X_{i+1} = 0, \forall i, i \geq 0$.
2. $|X_i - X_{i-1}| \leq \Delta, \forall i, i \geq 1$.
3. $E(X_i - X_{i-1}|X_{i-1}; 0 < X_{i-1} \leq R) \geq M, \forall i, i \geq 1$.

Let T *be the random variable defined as* $\min\{i \geq 0 | X_i = 0\}$. *Then*

$$Pr(T \leq t|X_0) \leq e^{-(MX_0/\Delta^2)} + te^{-(M(R-\Delta)/\Delta^2)}$$

Proof. Let λ denote $\frac{M}{\Delta^2} > 0$. First, using Markov's inequality:

$$\begin{aligned} Pr(T \leq t|X_0) &= Pr(X_t \leq 0|X_0) \\ &= Pr(\mathrm{e}^{-\lambda X_t} \geq 1|X_0) \\ &\leq E_{X_t}(\mathrm{e}^{-\lambda X_t}|X_0) \end{aligned}$$

We will show the following bound (to be proved shortly).

$$E_{X_i}(\mathrm{e}^{-\lambda X_i}|X_{i-1}) \leq \mathrm{e}^{-\lambda X_{i-1}} + \mathrm{e}^{-\lambda(R-\Delta)}, \qquad \forall i \geq 1 \tag{5.1}$$

Using this claim repeatedly below, the theorem follows.

$$\begin{aligned} E_{X_t}(\mathrm{e}^{-\lambda X_t}|X_0) &= E_{X_{t-1}}(E_{X_t}(\mathrm{e}^{-\lambda X_t}|X_{t-1})|X_0) \\ &\leq E_{X_{t-1}}(\mathrm{e}^{-\lambda X_{t-1}} + \mathrm{e}^{-\lambda(R-\Delta)}|X_0) \\ &\leq E_{X_{t-1}}(\mathrm{e}^{-\lambda X_{t-1}}|X_0) + \mathrm{e}^{-\lambda(R-\Delta)} \\ &\leq \mathrm{e}^{-\lambda X_0} + t\mathrm{e}^{-\lambda(R-\Delta)} \\ &= \mathrm{e}^{-(MX_0/\Delta^2)} + t\mathrm{e}^{-(M(R-\Delta)/\Delta^2)} \end{aligned}$$

The last equality follows by substituting for λ.

It remains to prove (5.1). The left-hand side of (5.1) is conditioned on a fixed value of X_{i-1}. When $X_{i-1} = 0$, $X_i = 0$ (from the first condition of the theorem), and the equation is trivially true. So assume that $X_{i-1} > 0$. This implies:

$$\begin{aligned} &E_{X_i}(\mathrm{e}^{-\lambda X_i}|X_{i-1}; X_{i-1} > 0) \\ &\quad\leq E_{X_i}(\mathrm{e}^{-\lambda X_i}|X_{i-1}; 0 < X_{i-1} \leq R) + E_{X_i}(\mathrm{e}^{-\lambda X_i}|X_{i-1}; X_{i-1} > R) \\ &\quad\leq E_{X_i}(\mathrm{e}^{-\lambda X_i}|X_{i-1}; 0 < X_{i-1} \leq R) + \mathrm{e}^{-\lambda(R-\Delta)} \\ &\quad= \mathrm{e}^{-\lambda X_{i-1}} E_{X_i}(\mathrm{e}^{-\lambda(X_i - X_{i-1})}|X_{i-1}; 0 < X_{i-1} \leq R) + \mathrm{e}^{-\lambda(R-\Delta)} \end{aligned}$$

The second inequality follows from condition (2) in the theorem and the fact that $\lambda > 0$. We prove below that $E_{X_i}(\mathrm{e}^{-\lambda(X_i - X_{i-1})}|X_{i-1}; 0 < X_{i-1} \leq R) \leq 1$, and (5.1) then follows.

We use the following fact. Suppose A and B are random variables having the same expectation. Further suppose that B takes on just two extreme values while A takes values in between these two extreme values. Let f be any convex function. Then $E(f(A)) \leq E(f(B))$.

Let $E_{X_i}(X_i - X_{i-1}|X_{i-1}; 0 < X_{i-1} \le R) = M_i$. Note that $M_i \ge M > 0$ [from condition (3) of the theorem]. Consider a random variable B which takes the value Δ with probability $\frac{1}{2}(1 + \frac{M_i}{\Delta})$ and $-\Delta$ with probability $\frac{1}{2}(1 - \frac{M_i}{\Delta})$; its expectation is M_i, which equals $E_{X_i}(X_i - X_{i-1}|X_{i-1}; 0 < X_{i-1} \le R)$. Then we obtain, as required:

$$\begin{aligned} E_{X_i}\left(e^{-\lambda(X_i - X_{i-1})}|X_{i-1}; 0 < X_{i-1} \le R\right) &\le E(e^{-\lambda B}) \\ &= \frac{1}{2}\left(1 - \frac{M_i}{\Delta}\right)e^{\lambda\Delta} + \frac{1}{2}\left(1 + \frac{M_i}{\Delta}\right)e^{-\lambda\Delta} \\ &\le \frac{1}{2}\left(1 - \frac{M}{\Delta}\right)e^{\lambda\Delta} + \frac{1}{2}\left(1 + \frac{M}{\Delta}\right)e^{-\lambda\Delta} \\ &\le \frac{1}{2}\left(e^{(-M/\Delta + \lambda\Delta)} + e^{(M/\Delta - \lambda\Delta)}\right) \\ &= 1 \end{aligned}$$

The last equality is by substituting for λ. The second inequality follows from the fact that the right-hand side can be rewritten as $\frac{1}{2}(e^{\lambda\Delta} + e^{-\lambda\Delta}) - \frac{M_i}{\Delta}(e^{\lambda\Delta} - e^{-\lambda\Delta}) \le \frac{1}{2}(e^{\lambda\Delta} + e^{-\lambda\Delta}) - \frac{M}{\Delta}(e^{\lambda\Delta} - e^{-\lambda\Delta})$. ■

6. BOUNDING THE COUPLING TIME

We now state the main theorem of the paper.

Theorem 6.1. *Consider any causal coupling process for the Jerrum–Sinclair chain. The probability that this process has coupled exceeds* $1 - \epsilon$ *only after time* $\Omega((1 - \epsilon - c'^n)e^{\Theta(n)})$, *where* $c' < 1$ *is as defined in the initial distribution specified in Section* 4. *In addition, the expected coupling time starting at any state in the coupling process having unequal components is* $\Omega(e^{\Theta(n)})$.

Proof. We shall use Theorem 5.1 by setting appropriate values for all the parameters.

Defining. X_i, Δ. X_0 is defined as the layer number of the starting vertex. Recall that $X_0 \ge n/4$ with probability $1 - c'^n$, for some constant $c' < 1$ (see the initial distribution in Section 4).

By virtue of Lemma 4.5, we may as well assume that the start vertex is in a *Bot*() set rather than a *Top*() set.

For $i > 0$, X_i is defined as follows. If $X_{i-1} = 0$ then $X_i = 0$. Otherwise, X_i is the layer number of the first state A reached in $\mathcal{M}$ with the following properties:

1. $A \notin L(X_{i-1})$.
2. A is in some *Bot*() set or in $L(0)$.

By Lemmas 4.1 and 4.5, we can set Δ to 8. Note that all X_i's are non-negative.

Let $p_i = \frac{\alpha n/2 - X_{i-1} - 2}{2m}$ and $q_i = \frac{4X_{i-1} + 2 + X_{i-1} + 3}{2m} = \frac{5(X_{i-1}+1)}{2m}$. We need to estimate $E(X_i - X_{i-1}|X_{i-1})$ for using Theorem 5.1. We shall need the following claim which lower bounds the probability that $X_i > X_{i-1}$.

Claim. $\Pr(X_i > X_{i-1}|X_{i-1}; X_{i-1} > 0) \geq \frac{p_i}{p_i+q_i}$

Proof. Note that $X_i \neq X_{i-1}$. The conditioning on X_{i-1} above indicates that X_i is defined by simulating $\mathcal{M}$ starting from a state in $Bot(X_{i-1})$ drawn according to some probability distribution, until a state in either $Bot(j)$, $j \neq X_{i-1}$ or $L(0)$ is reached; X_i is the layer number of the state reached.

By Lemma 4.5, it follows that $X_i < X_{i-1}$ only if the first vertex visited after leaving $Bot(X_{i-1})$ for the last time [note that the simulation could exit and revisit $Bot(X_{i-1})$ several times before the event defining X_i occurs] is either in $L(j)$, $j < i$ or in one of $Top(X_{i-1})$, $Top(X_{i-1}+1)$. By Lemma 4.2 and Lemma 4.4, this probability is at most q_i.

Also, if the first vertex visited on leaving $Bot(X_{i-1})$ for the last time is in $Bot(X_{i-1}+1) \cup \cdots \cup Bot(X_{i-1}+4)$, we would have $X_i > X_{i-1}$. By Lemma 4.3, this must happen with probability at least p_i.

The claim follows since $X_i \neq X_{i-1}$, by definition. ■

Defining M, R. Define $R = \beta n$ for some constant β satisfying $\frac{20(R+1)}{\alpha n/2-R-2} \leq \frac{1}{4}$. Using Lemma 6 and the fact that $\Delta = 8$, we bound $E_{X_i}(X_i - X_{i-1}|X_{i-1}; 0 < X_{i-1} \leq R)$ as follows.

$$\begin{aligned} E_{X_i}(X_i - X_{i-1}|X_{i-1}; 0 < X_{i-1} \leq R) &\geq \frac{p_i}{p_i+q_i} - \frac{8q_i}{p_i+q_i} \\ &\geq \frac{p_i - 8q_i}{2p_i} \\ &= \frac{1}{2} - \frac{4q_i}{p_i} \\ &= \frac{1}{2} - \frac{20(X_{i-1}+1)}{\alpha n/2 - X_{i-1} - 2} \\ &\geq \frac{1}{2} - \frac{20(R+1)}{\alpha n/2 - R - 2} \\ &\geq 1/4 \end{aligned}$$

So we set M to $\frac{1}{4}$.

Applying Theorem 5.1. With the above mentioned values of $\Delta = 8, M = \frac{1}{4}, R = \beta n$, and $X_0 \geq n/4$, the previous theorem yields the following for the coupling time T.

$$1 - \epsilon \leq \Pr(T \leq t_\epsilon | X_0) \leq c'^n + t_\epsilon e^{-\Theta(n)}$$

Here, the first term on the right is for the probability that $X_0 < n/4$, and the second term covers the case when $X \geq n/4$ and uses Theorem 5.1. It follows that $t_\epsilon \geq (1 - \epsilon - c'^n)e^{\Theta(n)}$ and the first part of the theorem follows, since t_ϵ is defined as the earliest instant at which the probability that the coupling time exceeds t_ϵ falls below ϵ.

The expected coupling time, $E(T|X_0)$, can be bounded by $\Omega(e^{\Theta(n)})$, even when $0 < X_0 = O(1)$ as follows:

$$\begin{aligned} E(T|X_0) &= \sum_{t\geq 0} \Pr(T > t|X_0) \\ &\geq \sum_{t\geq 0} \max\{0, (1 - e^{-(MX_0)/(\Delta^2)} - te^{-(M(R-\Delta)/\Delta^2)})\} \\ &\geq \sum_{t\geq 0} \max\{0, (1 - e^{-(M/\Delta^2)} - te^{-\Theta(n)})\} \\ &\geq \sum_{t\geq 0} \max\{0, (\Theta(1) - te^{-\Theta(n)})\} \\ &= \Omega(e^{\Theta(n)}) \end{aligned}$$

■

7. CONCLUSIONS

We have shown that causal coupling is not powerful enough to prove rapid mixing of the Jerrum–Sinclair chain on certain graphs. Curiously, our argument does not go through directly for Broder's original chain [3], due to the one-sided nature of its moves (the edge swapped in is always incident on the unmatched vertex in V_1, the left side of G). It would be interesting to see whether our result can be extended to Broder's chain as well. Another question is whether our result can be extended to a larger class of Markov chains; we leave this as an open problem.

ACKNOWLEDGMENTS

We thank Vivek Borkar for several very useful discussions related to this work. We thank Ravi Kannan for listening to early versions of the proof. We thank Alistair Sinclair and an anonymous referee for pointing out to us the existence of maximal noncausal couplings [12]. We also thank the above referee for informing us of [7], which deals with issues related to those addressed in this paper. We thank an anonymous referee for pointing us to [13, 26] and suggesting the use of the submartingale inequality; this has simplified the paper considerably in comparison to the preliminary version [2]. Finally, we thank several anonymous referees for numerous comments which have improved readability.

REFERENCES

[1] D. Aldous, Random walks on finite groups and rapidly mixing Markov chains, Séminaire de probabilités XVII 1981/82, Lecture Notes in Math., vol. 986, Springer-Verlag, New York, 1983, pp. 243–297.

[2] V.S. Anil Kumar and H. Ramesh, Markovian coupling vs. conductance for the Jerrum–Sinclair chain, Proc 40th Annual IEEE Symp Foundations of Computer Science, 1999, pp. 241–251.

[3] A.Z. Broder, How hard is it to marry at random? (On the approximation of the permanent), Proc 18th ACM Symp Theory of Computating, 1986, pp. 50–58. Erratum in Proc 20th ACM Symp Theory of Computing, 1988, p. 551.

[4] R. Bubley and M. Dyer, Path coupling: a technique for proving rapid mixing in Markov chains, Proc 38th Ann IEEE Symp Foundations of Computer Science, 1997, pp. 223–231.

[5] R. Bubley and M. Dyer, Faster random generation of linear extensions, Proc 9th ACM-SIAM Symp Discrete Algorithms, 1998, pp. 350–354.

[6] R. Bubley, M. Dyer, and M. Jerrum, An elementary procedure for sampling points in a convex body, Random Struct Alg 12 (1998), 213–235.

[7] K. Burdzy and W. Kendall, Efficient Markovian couplings: examples and counter examples, Research Report 331, University of Warwick, Dept. of Statistics, 1998.

[8] A. Czumaj, P. Kanarek, M. Kutylowski, and K. Lorys, Delayed path coupling and generating random permutations via distributed stochastic processes, Proc 10th ACM-SIAM Symp Discrete Algorithms, 1999, pp. 271–280.

[9] P. Diaconis, Group representations in probability and statistics, Institute of Mathematical Statistics, 1988.

[10] M. Dyer, A. Frieze, and R. Kannan, A random polynomial time algorithm for approximating the volume of convex bodies, J ACM, 38 (1991), 1–17.

[11] E.B. Dynkin and A.A. Yushkevich, Controlled Markov processes, Springer Verlag, Berlin, 1979.

[12] D. Griffeath, A maximal coupling for Markov chains, Z. Wahrscheinlichkeitstheorie Verw 31 (1975), 95–106.

[13] B. Hajek, Hitting time and occupation time bounds implied by drift analysis with applications, Adv Appl Probab 14 (1982), 502–525.

[14] M. Jerrum, A very simple algorithm for estimating the number of k-colourings of a low degree graph, Random Struct Alg 7 (1995), 157–165.

[15] M. Jerrum, Mathematical foundations of the Markov chain Monte Carlo method, Probabilistic methods for algorithmic discrete mathematics (M. Habib, C. McDiarmid, J. Ramirez-Alfonsin, B. Reed, eds.), Algo. Combin vol. 19, Springer Verlag, New York, 1998, pp. 116–165.

[16] M. Jerrum, An analysis of a Monte Carlo algorithm for estimating the permanent, Technical Report ECS-LFCS-91-164, Dept. of Computer Science, Univ. of Edinburg, 1991.

[17] M. Jerrum and A. Sinclair, Approximating the permanent, SIAM J Comput 18 (1989) 1149–1178.

[18] M. Jerrum and A. Sinclair, Polynomial time approximation algorithms for the Ising model, SIAM J Comput, 22 (1993), 1087–1116.

[19] R. Kannan, Markov chains and polynomial time algorithms, Proc 35th Ann IEEE Symp Foundations of Computer Science, 1994, pp. 656–671.

[20] T. Lindvall, Lectures on the coupling method, Wiley, New York, 1992.

[21] L. Lovasz and M. Simonovits, Random walks in a convex body and an improved volume algorithm, Random Struct Alg 4 (1993), 359–412.

[22] M. Luby and E. Vigoda, Approximately counting up to four, Proc 29th ACM Symp Theory of Computating, 1997, pp. 682–687.

[23] M. Mihail, On coupling and the approximation of the permanent, Inform Process Lett 30 (1989), 91–95.

[24] J.S. Rosenthal, Convergence rates for Markov chains, SIAM Rev 37 (1995), 387–405.

[25] A. Sinclair, Randomised algorithms for counting and generating combinatorial structures, Birkhäuser, Zurich, 1993.

[26] G. Sasaki and B. Hajek, The time complexity of maximum matching by simulated annealing, J ACM 35(2) (1988), 387–403.